Multiple introductions of Zika virus into the United States revealed through genomic epidemiology

Nathan D Grubaugh[1*], Jason T Ladner[2,*], Moritz UG Kraemer[3,*], Gytis Dudas[4,*], Amanda L Tan[5,*], Karthik Gangavarapu[1*], Michael R Wiley[2,*], Stephen White[6,*], Julien Thézé[3,*], Diogo M Magnani[7], Karla Prieto[2], Daniel Reyes[2], Andrea Bingham[8], Lauren M Paul[5], Refugio Robles-Sikisaka[1], Glenn Oliveira[9], Darryl Pronty[6], Hayden C Metsky[10], Mary L Baniecki[10], Kayla G Barnes[10], Bridget Chak[10], Catherine A Freije[10], Adrianne Gladden-Young[10], Andreas Gnirke[10], Cynthia Luo[10], Bronwyn MacInnis[10], Christian B Matranga[10], Danny J Park[10], James Qu[10], Stephen F Schaffner[10], Christopher Tomkins-Tinch[10], Kendra L West[10], Sarah M Winnicki[10], Shirlee Wohl[10], Nathan L Yozwiak[10], Joshua Quick[11], Joseph R Fauver[12], Kamran Khan[13], Shannon E Brent[13], Robert C. Reiner Jr.[14], Paola N Lichtenberger[7], Michael Ricciardi[7], Varian K Bailey[7], David I Watkins[7], Marshall R Cone[15], Edgar W Kopp IV[15], Kelly N Hogan[15], Andrew C Cannons[15], Robert F Garry[16], Nicholas J Loman[11], Nuno R Faria[3], Mario C Porcelli[17], Chalmers Vasquez[17], Derek AT Cummings[18], Elyse Nagle[2], Danielle Stanek[8], Andrew Rambaut[19,20], Mariano Sanchez-Lockhart[2], Pardis C Sabeti[10,#], Leah D Gillis[6,#], Scott F Michael[5,#], Trevor Bedford[4,#], Oliver G Pybus[3,#], Sharon Isern[5,#], Gustavo Palacios[2,#,$], Kristian G Andersen[1,9,21,#,$]

[1]Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, CA 92037, USA
[2]Center for Genome Sciences, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, MD 21702, USA
[3]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK
[4]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[5]Department of Biological Sciences, College of Arts and Sciences, Florida Gulf Coast University, Fort Myers, FL 33965, USA
[6]Bureau of Public Health Laboratories, Division of Disease Control and Health Protection, Florida Department of Health, Miami, FL 33125, USA
[7]Department of Pathology, University of Miami Miller School of Medicine, Miami, FL 33136, USA
[8]Bureau of Epidemiology, Division of Disease Control and Health Protection, Florida Department of Health, Tallahassee, FL 32399, USA
[9]Scripps Translational Science Institute, La Jolla, CA 92037, USA
[10]The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[11]Institute of Microbiology and Infection, University of Birmingham, Birmingham B15 2TT, UK
[12]Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523, USA
[13]Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON M5B 1T8, Canada
[14]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA 98121, USA
[15]Bureau of Public Health Laboratories, Division of Disease Control and Health Protection, Florida Department of Health, Tampa, FL 33612, USA
[16]Department of Microbiology and Immunology, Tulane University School of Medicine, New Orleans, LA 70112, USA
[17]Miami-Dade County Mosquito Control District, Miami, FL 33178 USA
[18]Department of Biology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32610, USA
[19]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, UK
[20]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA
[21]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[*] = co-first
[#] = co-senior
[$] = co-corresponding

Zika virus (ZIKV) is currently causing an unprecedented pandemic linked to severe congenital syndromes[1,2]. In July 2016, mosquito-borne ZIKV transmission was first reported in the continental United States and since then, hundreds of locally acquired infections have been described[3]. To gain insights into the timing, source, and likely route(s) of introduction into the United States, we tracked the virus from its first detection in Miami, Florida by direct sequencing of ZIKV genomes from infected patients and *Aedes aegypti* mosquitoes. We detected at least four distinct ZIKV introductions and estimate that 11-52 introductions contributed to the outbreak in Florida. Furthermore, our data suggests that ZIKV transmission likely started in the spring of 2016 - several months before initial detection. By analyzing epidemiological, surveillance, and genetic data, we discovered that several spatially distinct ZIKV transmission zones were likely portions of the same outbreak, rather than isolated events. Our analyses show that most introductions are linked to the Caribbean, which is supported by the high incidence rates and traffic, especially via cruises, from the region into Miami. By comparing mosquito abundance and travel capacity across the United States, we find that southern Florida is especially vulnerable to ZIKV introductions and at risk of repeat occurrences. By tracking the virus from its initial introduction into the United States, we provide a deeper understanding of how ZIKV initiates and sustains transmission in new regions.

Local ZIKV transmission was first reported in Brazil in May 2015[4], though it was likely introduced to the Americas 1-2 years prior to its detection[5]. By January 2016, the ZIKV epidemic had spread to several South and Central American countries and most Caribbean islands[6]. Like dengue virus (DENV) and chikungunya virus (CHIKV), ZIKV is vectored primarily by *Aedes* mosquitoes[7–10]. The establishment of the peridomestic mosquito *Ae. aegypti* in the Americas[11] has facilitated DENV, CHIKV, and now likely ZIKV[12] to become endemic in this region. In the United States, transient outbreaks of DENV and CHIKV have been reported in regions of Texas and Florida[13–17] with abundant seasonal *Ae. aegypti* populations[11,18].

In July 2016, the first locally-acquired ZIKV cases in the continental United States were detected in Miami, Florida[3]. The outbreak, which generated 256 confirmed locally acquired ZIKV infections in 2016, was likely initiated by one or more of the many travel-associated infections in Florida (Fig. 1a). While transmission was confirmed across four counties in Florida (Fig. 1b), the ZIKV outbreak was the most intense in Miami-Dade County (244 infections). Although the location of exposure could not be determined in all cases, at least 114 (45%) of the infections were likely acquired in one of three distinct transmission zones: Wynwood, Miami Beach, and Little River (Fig. 1c-d).

We obtained mosquito surveillance data from Miami-Dade County's Mosquito Control Division (Fig. 1c-d and Extended Data Fig. 1). Of 24,351 mosquitoes (sorted into 2,596 pools) collected throughout Miami-Dade County from June to November, 2016, 99.8% were *Ae. aegypti*. Eight of these pools tested positive for ZIKV RNA, all of which were *Ae. aegypti* collected in Miami Beach (Fig. 1c). From these pools, we estimated that ~1 out of 1,600 *Ae. aegypti* mosquitoes were infected (0.061%, 95% CI: 0.028-0.115%, Extended Data Fig. 1a), which is similar to some reported *Ae. aegypti* infection rates during DENV and CHIKV outbreaks[19]. Even though ZIKV-infected mosquitoes were not detected in transmission zones outside Miami Beach (Fig. 1c), we found that the number of human ZIKV cases correlated strongly with *Ae. aegypti* abundance within each of the defined transmission zones (Spearman r = 0.61, Fig. 1d, Extended Data Fig. 1e). This suggests that the primary mode of transmission throughout Miami was via *Ae. aegypti* mosquitoes and that mosquito abundance changes directly impacted human infection rates. Therefore, the application of insecticides[3], which we found to suppress local mosquito populations during the periods of intensive usage (Extended Data Fig. 1c), likely contributed to the clearance of ZIKV within the Miami area.

To gain molecular insights into the Florida outbreak, we sequenced 39 ZIKV genomes directly from clinical and mosquito samples without cell culture enrichment[20] (Supplementary Table 1). Our Florida

ZIKV dataset (36 genomes) included 29 genomes from patients with locally-acquired infections (Fig. 1d) and 7 from *Ae. aegypti* mosquito pools (Fig. 1c). We also sequenced 3 ZIKV genomes from travel-associated cases diagnosed in Florida. Our dataset included cases associated with all three transmission zones in Miami (Fig. 1d) and represented ~11% of all confirmed locally-acquired cases in Florida. We made all sequence data openly available at NCBI (BioProjects PRJNA342539, PRJNA356429) and other online resources shortly after data generation.

We reconstructed phylogenetic trees from our 39 ZIKV genomes along with 65 published genomes from other affected regions to infer viral genealogies (Fig. 2, Extended Data Fig. 2 and 3). We found that the Florida ZIKV genomes formed four distinct lineages (F1 - F4, Fig. 2a), three of which (F1 - F3) belong to the same major clade (clade A, Fig. 2a). We only sampled a single human case each from the F3 and F4 lineages, consistent with limited transmission (Fig. 2a). The other two Florida lineages (F1 and F2) were derived from several human and mosquito samples from Miami-Dade County (Fig. 2b).

We estimated the number of introductions responsible for the locally acquired cases observed in our dataset using time-structured phylogenies[21]. Lineage F4 resulted from an independent introduction based on phylogenetic placement within a distinct major clade (Fig. 2a). We estimated the time of the most recent common ancestor (tMRCA) to be in the summer of 2015 for the two well-supported nodes (A and B) linking lineages F1, F2, and F3 (Fig. 2a). The tMRCA estimates were robust across a range of different molecular clock and coalescent models (Extended Data Table 1). This suggests that while F1-F3 belong to the same clade, any less than three distinct introductions leading to these lineages would likely have required undetected transmission of ZIKV in Florida for approximately one year (Fig. 2a).

To estimate the likelihood of a single transmission chain persisting for over a year, we modeled epidemic spread under different assumptions of the basic reproductive number (R_0). We found that R_0 of approximately 0.6 was most consistent with locally acquired and travel associated case counts in Miami-Dade County (Extended Data Fig. 4). With R_0 of 0.6, we determined that the probability of a single transmission chain persisting for over a year in Florida is extremely low (< 0.0001, Fig. 2c). This is especially true considering that *Ae. aegypti* abundance is low in Florida during the winter months (Extended Data Fig. 1f). Given the low probability of long term persistence, we expect that each sampled ZIKV lineage in Florida (F1-F4) was likely the result of a separate introduction. Differences in surveillance practises across areas, combined with the high number of travel-associated cases in Florida (Fig. 1a), likely mean that unsampled ZIKV introductions have also contributed to the size of the outbreak. Therefore, to estimate the total number of underlying ZIKV introductions, we modeled epidemic dynamics resulting in 244 locally acquired cases within Miami-Dade County, and found that with R_0 of 0.6 we expect 31 (95% CI 11–52) separate introductions to have driven the Miami-Dade outbreak (Fig. 2d). The majority of these introductions would likely have generated a single secondary case that was undetected in the genetic sampling (Extended Data Fig. 4a).

The two main ZIKV lineages, F1 and F2, included the vast majority of genomes we sampled in Florida (92%, Fig. 2a). We investigated the timing of introduction of these lineages, and found that the probability densities for the tMRCAs of both F1 and F2 were centered around March-April, 2016 (Fig. 2b, Extended Data Fig. 3, 95% Bayesian credible intervals (BCI): January-May, 2016). This finding suggests that ZIKV transmission in Florida started at least two months prior to its first detection in July 2016 (Fig. 1a).

To investigate if the ZIKV transmission zones within Miami were connected, we analyzed our genomic data together with case investigation data from the Florida Department of Health (DOH; Supplementary Table 1). While spatially distinct, the ZIKV transmission zones occurred within ~3 miles of each other (Fig. 1c) and we found that the ZIKV infections associated with each zone overlapped temporally (Fig. 1d). Our 22 ZIKV genomes with zone assignments all belonged to lineages F1 and F2, but neither of these lineages were confined to a single transmission zone (Fig. 2b). In fact, we detected both F1 and F2

lineage viruses from *Ae. aegypti* collected from the same trap 26 days apart (mosquitoes 5 and 8, Fig. 2b). The temporal overlap of ZIKV cases and the detection of genetically similar viruses across zones suggest that the ZIKV transmission zones within Miami were part of the same outbreak, rather than independent events.

ZIKV transmission is not likely to persist through the winter in Florida[22] (Fig. 2c, Extended Data Fig. 1f); therefore, determining the sources and routes of ZIKV introductions could help to avert future outbreaks. We found that lineages F1-F3 clustered with ZIKV genomes sequenced from the Dominican Republic and Guadeloupe (Extended Data Figs 2 and 3), suggesting a Caribbean source for these viruses (Fig. 2). In contrast, the F4 lineage clustered with genomes from Central America (Extended Data Figs 2 and 3). Overall, our findings suggest that while ZIKV outbreaks occurred throughout the Americas, the Caribbean islands appear to have been the main source of ZIKV that resulted in local transmission in Florida. Because of the severe undersampling of ZIKV genomes from across the Americas, however, we cannot rule out other source areas or panmixis of the ZIKV population in the region. Similarly, even though we found that the Florida ZIKV genomes clustered together with sequences from the Dominican Republic, our results should not be interpreted to indicate that ZIKV entered Florida from this country.

We investigated ZIKV infection rates and travel patterns to corroborate our phylogenetic evidence for Caribbean introductions into Florida. We found that the Caribbean islands bore the highest ZIKV incidence rates (Fig. 3a, Extended Data Fig. 5, Supplementary Table 2), despite Brazil and Colombia reporting the highest absolute number of ZIKV cases around the likely time of most introductions into Florida (January-June, 2016). Over the same time period, travelers from the Caribbean accounted for 56% (~3 million) of the total traffic into Miami, with the vast majority (~2.4 million) arriving via cruise ships (Fig. 3b, Extended Data Fig. 6, Supplementary Table 2). Combining the infection rates with travel capacities, we estimated that ~60-70% of ZIKV infected travelers entering Miami arrived from the Caribbean (Fig. 3c and Extended Data Fig. 7a). In support of our estimates, we found that the observed number of travel-associated ZIKV cases correlated strongly with the expected number of importations from the Caribbean (Spearman r = 0.8, Fig. 3d, Extended Data Fig. 7b) and 45% of the people infected abroad reported recent travel to the Caribbean (Fig. 3e). Notably, 23% of the infected travelers diagnosed in Miami visited Nicaragua (Central America). Because there are no ZIKV genomes sampled from this country, we cannot determine the relationship between the outbreaks in Nicaragua and Florida. Taken together, these findings suggest that a high incidence of ZIKV in the Caribbean, combined with frequent cruise ship travel from the region, could have played key roles in the establishment of ZIKV transmission in Florida in the spring of 2016. These findings, however, should not be interpreted to indicate that cruise ships themselves are risk factors for human ZIKV infection, but only that they serve as a major mode of transportation from areas with active transmission.

The vast majority of the ZIKV outbreak in the United States occurred in Miami-Dade County (95% of reported local cases, Fig. 1b). To investigate if ZIKV transmission was more likely to become established in Miami than in other parts of Florida, we analyzed *Ae. aegypti* abundance and incoming passenger traffic across all Florida ports. During January to April 2016 when the earliest ZIKV introductions likely occurred (Fig. 2b), we estimated that mosquito abundance was higher near Miami compared to other regions of Florida (Fig. 3f, Extended Data Fig. 1f); however, by June, all Florida counties were predicted to support relatively high *Ae. aegypti* populations (Fig. 3f). In addition, we found that ports in Miami and nearby Fort Lauderdale received ~72% of the total flight and cruise traffic entering Florida (Fig. 3f, Extended Data 8a). Therefore, we expect a higher likelihood of ZIKV introductions and transmission in the greater Miami area, due to the high volume of incoming traffic, dense human population, and high early-season *Ae. aegypti* abundance. These findings could also explain why United States-based outbreaks of DENV and CHIKV have mostly occurred in Miami and surrounding areas[16,17].

We analyzed travel data from passengers arriving from ZIKV-afflicted regions and seasonal *Ae. aegypti* abundances[18] to determine if other regions in the continental United States were at risk for ZIKV

outbreaks. We found that Florida received 8 times more cruise ship (Extended Data Fig. 8b) and 1.5 times more flight passengers[23] per month than any other state. In addition, southern Florida supports high *Ae. aegypti* populations most of the year[11,18] (Extended Data Fig. 8c). This helps explain why Miami is one of the regions in the United States most at risk for *Ae. aegypti*-borne viruses, including ZIKV, DENV, and CHIKV[24]. Outside of Florida, major port cities such as Houston, New York, and New Orleans received a high proportion of international traffic[23] (Extended Data Fig. 8b), support seasonal *Ae. aegypti* populations[18] (Extended Data Fig. 8c), and therefore may be at risk of transient ZIKV outbreaks. In the last 20 years, however, the only area in the continental United States outside of Florida to report *Ae. aegypti*-borne virus transmission is along the United States-Mexico border near Brownsville, TX[25]. This area also reported local ZIKV transmission in 2016[26], which, as suggested for the 2005 DENV outbreak[17], may have been initiated by land travel from Mexico[23]. Therefore, Caribbean cruises may only be an important mode of ZIKV introductions in Florida (Fig. 3, Extended Data Fig. 8c).

The extent of ZIKV transmission in Florida was unprecedented. There were more reported ZIKV cases in 2016 (256) than all DENV cases observed since 2009 (135)[13]. Given that the majority of ZIKV infections are asymptomatic[2,27], these case numbers are likely underestimated. Despite this, we predict that R_0 was less than 1 and multiple introductions were probably necessary to fuel an outbreak of this magnitude[22]. Human air travel is often implicated in the spread of mosquito-borne viruses[28], but here we suggest that cruise ships traveling from Caribbean islands with intense ZIKV transmission may have provided a substantial supply of the ZIKV infections that contributed to the Florida outbreak. Our virus genetic data also link three of the four observed ZIKV introductions to the Caribbean islands; though because of undersampling in the region, we cannot pinpoint exact sources. ZIKV is likely to become endemic to the Americas[12], and because Miami supports relatively large *Ae. aegypti* populations[11,18], is densely populated, and is a hub for international travel[23], we expect that future ZIKV introductions into Florida will occur. However, the magnitude of future outbreaks is likely to be limited in the absence of repeated introduction events.

Acknowledgements

The authors thank the Florida DOH and Miami-Dade Mosquito Control for providing data regarding the ZIKV outbreak in Florida, Chet Moore, Barry Alto, Sophie Taylor, and XXX for discussions, and Andrew Monaghan for providing details about the *Ae. aegypti* abundance data from United States port cities.

N.D.G. is supported by a National Institutes of Health (NIH) training grant 5T32AI007244-33.

G.D. is supported by the Mahan Postdoctoral Fellowship from the Computational Biology Program at Fred Hutchinson Cancer Research Center.

D.A.T.C. was supported by the US NIH MIDAS program (U54-GM088491) as well as Cooperative Agreement U01CK000510 funded by the CDC and Prevention.

A.R. was supported by the European Union Seventh Framework Programme [FP7/2007-2013] under Grant Agreement 278433-PREDEMICS and ERC Grant agreement 260864 and Horizon 2020 research and innovation program Grant Agreement 643476-COMPARE.

T.B. is a Pew Biomedical Scholar and his work is supported by NIH award R35 GM119774-01.

K.G.A. is a Pew Biomedical Scholar, and his work is supported by an NIH National Center for Advancing Translational Studies Clinical and Translational Science Award UL1TR001114, and National Institute of Allergy and Infectious Diseases (NIAID) contract HHSN272201400048C.

The USAMRIID Center for Genome Sciences is supported by the Defense Threat Reduction Agency Joint Science and Technology Office (DTRA JSTO). ZIKV sequencing at USAMRIID was supported by the Defense Advanced Research Projects Agency (DARPA).

The content of this publication does not necessarily reflect the views or policies of the US Army, the CDC and Prevention, the US Department of Health and Human Services, or the Florida Department of Health.

Author Contributions

Designed the experiments: K.G.A., G.P., S.I., O.G.P., T.B., S.F.M., L.D.G., P.C.S., N.D.G., J.T.L., G.D., M.U.G.K., and D.A.T.C.

Collected samples: A.L.T., S.W., D.M.M., A.B., L.M.P., D.P., P.N.L., M.R., V.K.B., D.I.W., M.R.C., E.W.K., K.N.H., A.C.C., M.C.P., C.V., D.S., L.D.G., S.F.M., and S.I.

Performed the sequencing: N.D.G., J.T.L., M.W.R., K.P., D.R., R.R.-S., G.O., and E.N.

Provided data, critical reagents, or protocols: N.D.G., J.T.L., G.D., M.U.G.K., K.G., M.R.W., R.R.-S., G.O., H.C.M., M.L.B., K.G.B., B.C., C.A.F., A.G.-Y., A.G., C.L., B.M., C.B.M., D.J.P., J.Q., S.F.S., C.T.-T., K.L.M., S.M.W., S.W., N.L.Y., J.Q., J.R.F., K.K., S.E.B., R.F.G., N.J.L., M.C.P., C.V., P.C.S., S.F.M., and S.I.

Analyzed the data: N.D.G., J.T.L., G.D., M.U.G.K., K.G., J.T., J.R.F., R.C.R., N.R.F., D.A.T.C., A.K., M.S.-L., T.B., O.G.P., and K.G.A.

Wrote the manuscript: N.D.G., J.T.L., and K.G.A.

Edited the manuscript: G.D., M.U.G.K., J.T., S.F.S., A.R., and T.B.

All authors read and approved the contents of the manuscript.

Author Information

Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.G.A (andersen@scripps.edu) or G.P. (gustavo.f.palacios.ctr@mail.mil).

Methods

Ethical statement

This study has been evaluated and approved by Institutional Review Boards (IRB) at The Scripps Research Institute (TSRI). The study was reviewed by the Florida DOH Human Research Protection Program and the US Army Medical Research Institute of Infectious Diseases (USAMRIID) Office of Human Use and Ethics and determined not to require IRB review.

Florida Zika virus case data

Weekly reports of international travel-associated Zika fever cases in Florida and ZIKV infected cases acquired in Florida were obtained from the Florida DOH mosquito-borne disease surveillance system[13]. Dates of symptom onset from the Miami transmission zones (*i.e.* Wynwood, Miami Beach, and Little River) determined by the Florida DOH investigation process were obtained from the ZIKV resource website[29] and daily updates[30]. International travel-associated ZIKV case counts in the United States (outside of Florida) were obtained from The Centers for Disease Control and Prevention (CDC)[31]; due to some reporting discrepancies, however, the travel-associated case numbers for Florida were obtained from the Florida DOH and not the CDC.

Clinical sample collection

Clinical samples from locally-acquired ZIKV infections were collected from June 22 to October 11, 2016. The Florida DOH identified persons with compatible illness and clinical samples were shipped to the Bureau of Public Health Laboratories for confirmation by qRT-PCR and antibody tests following interim guidelines[3,32–34]. Clinical specimens (whole blood, serum, saliva, or urine) submitted for analysis were refrigerated or frozen at ≤−70°C until RNA was extracted. RNA was extracted using the RNAeasy kit (QIAGEN), MagMAX for Microarrays Total RNA Isolation Kit (Ambion), or MagNA Pure LC 2.0 or 96 Systems (Roche Diagnostics). Purified RNA was eluted into 50-100 µL of a low-salt buffer, immediately frozen at ≤ −70°C, and transported on dry ice. The Florida DOH also provided investigation data for these samples, including symptoms onset dates and, when available, assignments to the zone where transmission likely occurred (Supplementary Table 1).

Mosquito collection and entomological data analysis

24,351 *Ae. aegypti* and *Ae. albopictus* mosquitoes (sorted into 2,596 pools) were collected throughout Miami-Dade County during June to November, 2016 using BG-Sentinel mosquito traps (Biogents AG). Up to 50 mosquitoes of the same species and sex were pooled per trap. The pooled mosquitoes were stored in RNAlater (Invitrogen), RNA was extracted using either the RNAeasy kit (QIAGEN) or MagMAX for Microarrays Total RNA Isolation Kit (Ambion), and ZIKV RNA was detected by qRT-PCR targeting the envelope protein coding region[34] or the Trioplex qRT-PCR kit[35]. ZIKV infection rates were calculated per 1,000 female *Ae. aegypti* mosquitoes using the bias-corrected maximum likelihood estimate (MLE)[36]. Vector index was calculated by multiplying the *Ae. aegypti* abundance per trap night with the estimated proportion of infected mosquitoes[37]. MLE infection rate and vector index could only be calculated from Miami Beach because it was the only region with mosquitoes confirmed to contain ZIKV RNA. *Ae. aegypti* abundance, MLE infection rate, and vector were calculated using a sliding window of three days with a step of one day (i.e. mean and 95% CI of day before and after) to smooth some of the natural variability of daily entomological measures. Days of insecticide usage by the Miami-Dade Mosquito Control were inferred from the zone-specific ZIKV activities timelines published by the Florida DOH[29]. Daily minimum temperatures for Miami were acquired from NOAA's National Centers for Environmental Information[38].

Relative monthly *Ae. aegypti* abundance

For the purpose of this study we used previous modelling outputs from Kraemer et al.[11] reapplied the statistical relationships between mosquito presence and its environmental correlated to monthly covariate data to generate seasonally varying maps of probability of occurrence at 5 km × 5 km global grids[39].

Following Hwang et al.[40] we use a simple mathematical formula to transform the probability of detection maps into mosquito abundance maps. In order to do so, assume P (Y=1) where Y is a binary variable (presence/absence). Using a Poisson distribution X() to govern the abundance of mosquitoes, the probability of not observing any mosquitoes can be related to the probability of absence as: P(X=0)=P(Y=0). Now we use the following transformation to generate abundance estimates per county in Florida:

Zika virus quantification

ZIKV genome equivalents (GE) were quantified by qRT-PCR. At TSRI, ZIKV qRT-PCR was performed as follows: ZIKV RNA standards were transcribed from the ZIKV NS5 region (8651-9498 nt) using the T7 forward primer (5' - TAA TAC GAC TCA CTA TAG GGA GA + TCA GGC TCC TGT CAA AAC CC - 3'), reverse primer (5' - AGT GAC AAC TTG TCC GCT CC - 3'), and the T7 Megascript kit (Ambion). For qRT-PCR, primers and a probe targeting the NS5 region (9014-9123 nt) were designed using the ZIKV isolate PRVABC59 (GenBank: KU501215): forward primer (5'- AGT GCC AGA GCT GTG TGT AC - 3'), reverse primer (5' - TCT AGC CCC TAG CCA CAT GT - 3'), and FAM-fluorescent probe (5' - GGC AGC CGC GCC ATC TGG T - 3'). The qRT-PCR assays were performed in 25 µl reactions using the iScript One-step RT-PCR Kit for probes (Bio-Rad Laboratories Inc.) and 2 µl of sample RNA. Amplification was performed at 50°C for 20 min, 95°C for 3 min, and 40 cycles of 95°C for 10 s and 57°C for 10 s. Fluorescence was read at the end of the 57°C annealing-extension step. 10-fold dilutions of the ZIKV RNA transcripts (2 µl/reaction) were used to create a standard curve for quantification of ZIKV GE/µl of RNA.

ZIKV GE were quantified at USAMRIID using the University of Bonn ZIKV envelope protein (Bonn E) qRT-PCR assay[41]. RNA standards were transcribed using an amplicon generated from a ZIKV plasmid containing T7 promoter at the start of the 5'UTR region. The plasmid was designed using the ZIKV isolate Beh819015 (GenBank: KU365778) and the amplicon included nts 1-4348, which covers the 5'UTR, C, prM, M, E, NS1, and NS2 regions. The qRT-PCR assays were performed in 25 µl reactions using the SuperScript III platinum One-step qRT-PCR Kit (ThermoFisher) and 2 µl of sample RNA was used. Amplification was performed following conditions as previously described[41]. 10-fold dilutions of the ZIKV RNA transcripts (5 µl/reaction) were used to create a standard curve for quantification of ZIKV GE/µl of RNA.

Amplicon-based Zika virus sequencing

ZIKV sequencing at TSRI was performed using an amplicon-based approach, as described[20]. Briefly, cDNA was reverse transcribed from 5 µl of RNA using SuperScript IV (Invitrogen). ZIKV cDNA (2.5 µl/reaction) was amplified in 35 × 400 bp fragments from two multiplexed PCR reactions using Q5 DNA High-fidelity Polymerase (New England Biolabs). The amplified ZIKV cDNA fragments (50 ng) were prepared for sequencing using the Kapa Hyper prep kit (Kapa Biosystems) and SureSelect XT2 indexes (Agilent). Agencourt AMPure XP beads (Beckman Coulter) were used for all purification steps. Paired-end 251 nt reads were generated on the MiSeq using the V2 500 cycle or V3 600 cycle kit (Illumina).

Trimmomatic was used to remove primer sequences (first 22 nt from the 5' end of the reads) and bases at both ends with Phred quality score <20[42]. The reads were then aligned to the complete genome of a ZIKV isolate from the Dominican Republic, 2016 (GenBank: KU853012) using Novoalign v3.04.04 (www.novocraft.com). Samtools was used to sort the aligned BAM files and to generate alignment statistics[43]. Snakemake was used as the workflow management system[44]. The code and reference indexes

for the pipeline can be found at https://github.com/andersen-lab/zika-pipeline. ZIKV-aligned reads were visually inspected using Geneious v9.1.5[45] before generating consensus sequences. A minimum of 3x read-depth coverage, in support of the consensus, was required to make a call.

The consensus ZIKV sequences from FL01M and FL03M generated by sequencing 35 × 400 bp amplicons on the MiSeq were validated using the following approaches: 1) sequencing the 35 × 400 bp amplicons on the Ion S5 platform, 2) sequencing amplicons generated using an Ion AmpliSeq panel customly targeted towards ZIKV on the Ion S5 platform, and 3) sequencing 5 × 2,150-2,400 bp ZIKV amplicons on the MiSeq. For Ion library preparation, cDNA was synthesized using the SuperScript VILO kit (ThermoFisher). ThermoFisher designed 875 custom ZIKV primers to produce 75 amplicons of ~200 bp in two PCR reactions for use with their Ion AmpliSeq Library Kit 2.0. The reagent FuPa was used to digest the modified primer sequences after amplification. The DNA templates were loaded onto Ion 520 chips using the Ion Chef and sequenced one the Ion S5 with the 200 bp output (ThermoFisher). The 35 × 400 bp amplicons generated for the MiSeq as described above were introduced into the Ion workflow using the Ion AmpliSeq Library Kit 2.0, but without fragmentation. Primers to amplify 2,150-2,400 bp ZIKV fragments were kindly provided by Dawn Dudly, Shelby O'Connor, and Dane Gellerup (AIDS Vaccine Research Laboratory, University of Wisconsin, Madison). Each fragment was amplified individually by PCR using the cDNA generated above, Q5 DNA High-fidelity Polymerase, and the following thermocycle conditions: 55 °C for 30 m, 94 and 68 °C for 3.5 m, 6 Agencourt AMPure XP beads, sheared to 300 to 400 nt fragments using the Covaris S2 sonicator, indexed and prepared for sequencing as described above, and sequenced using the MiSeq V2 500 cycle kit (paired-end 251 nt reads). Compared to the consensus sequences generated using 35 × 400 bp amplicons on the MiSeq, there were no consensus-level mismatches in the CDS using any of the other three approaches (Extended Data Table 2). There were, however, some mismatches in the 5' and 3' untranslated regions (where the genomic RNA is heavily structured), likely a result of PCR bias and decreased coverage depth.

Enrichment-based Zika virus sequencing

ZIKV sequencing at USAMRIID was performed using a targeted enrichment approach. Sequencing libraries were prepared using the TruSeq RNA Access Library Prep kit (Illumina) with custom ZIKV probes. Our probe set included 866 unique probes each of which was 80 nt in length (Supplementary Table 3). The probes were designed to cover the entire ZIKV genome and to encompass the genetic diversity present on GenBank on 01/14/2016. In total, 26 ZIKV sequences were used during probe design (Supplementary Table 3). Extracted RNA was fragmented at 94˚C for 0-60 seconds and each sample was enriched separately using a quarter of the reagents specified in the manufacturer's protocol. Samples were barcoded, pooled and sequenced using the MiSeq Reagent kit v3 (Illumina) on an Illumina MiSeq with a minimum of 2 × 151 bp reads. Dual indexing, with no overlapping indices, was used.

The random hexamer associated with read one and the Illumina adaptors were removed from the sequencing reads using Cutadapt v1.9.dev1[46], and low-quality reads/bases were filtered using Prinseq-lite v0.20.3[47]. Reads were aligned to a reference genome (KX197192.1) using Bowtie2 v2.0.6[48], duplicates were removed with Picard (http://broadinstitute.github.io/picard), and a new consensus was generated using a combination of Samtools v0.1.18[43] and custom scripts. Only bases with Phred quality score ≥20 were utilized in consensus calling, and a minimum of 3X read-depth coverage, in support of the consensus, was required to make a call; positions lacking this depth of coverage were treated as missing (*i.e.* called as ''N'').

Phylogenetic analyses

All published and available complete ZIKV genomes of the Asian genotype from the Pacific and the Americas were retrieved from GenBank public database as of December 2016. Sequences (n=65) were aligned together with ZIKV genomes generated in this study (n=39) using MAFFT[49] and manual curation. The multiple alignment contained 104 ZIKV sequences collected between 2013 and 2016, from the Pacific (American Samoa, French Polynesia, and Tonga), Brazil, other South and Central Americas (Guatemala, Mexico, Suriname, and Venezuela), the Caribbean (Dominican Republic, Guadeloupe, Haiti, Martinique, and Puerto Rico), and the United States.

In order to determine the temporal signal of the sequence dataset, a maximum likelihood phylogeny was first reconstructed with PhyML[50] using the Hasegawa-Kishino-Yano (HKY) nucleotide substitution model[51] and gamma distributed rates amongst sites[52], which was identified as the best fitting model for ML inference by jModelTest2[53]. Then, a correlation between root-to-tip genetic divergence and date of sampling was conducted in TempEst[54].

Bayesian phylogenetic analyses were performed using BEAST v.1.8.4[21] to infer time-structured phylogenies. We used a SDR06 nucleotide substitution model[55] with a non-informative continuous time Markov chain reference prior[56] on the molecular clock rate. Replicate analyses using multiple combinations of molecular clock and coalescent demographic models were explored to select the best fitting model by marginal likelihood comparison using path-sampling and stepping-stone estimation approaches[57–59] (Extended Data Table 3). The best fit model was a strict molecular clock along with a Bayesian skyline demographic model[60]. All the Bayesian analyses were run for 30 million Markov chain Monte Carlo steps, sampling parameters and trees every 3000 generations.

An initial phylogeny, rooted on lineages previously sampled in French Polynesia (Extended Data Fig. 2a) resulted in a strong correlation between ZIKV genetic divergence and sampling time (Extended Data Fig. 2b).

Expected number and distribution of local cases from Zika virus importations

We used branching process theory[61,62] to generate the offspring distribution (subsequent local cases) that is expected from a single introduction. The offspring distribution is modelled with a negative binomial distribution with mean R_0 and over-dispersion parameter k. The total number of cases, L, that is caused by a single importation (including the index case) after an infinite time[63] has the following form:

$$L = \frac{\Gamma(k\,j+j-1)}{\Gamma(k\,j)\,\Gamma(j+1)} \frac{\left(\frac{R_0}{k}\right)^{j-1}}{\left(1+\frac{R_0}{k}\right)^{k\,j+j-1}}$$

The parameter k represents the variation in the number of secondary cases generated by each case of ZIKV[61]. In the case of vector borne diseases, local heterogeneity is high due to a variety of factors such as mosquito population abundance, human to mosquito interaction, and control interventions[64,65]. Here, we assume high heterogeneity (k=0.1) following previous estimates for vector borne diseases[62], though we conducted sensitivity analyses with more homogeneous transmission (k=1). This distribution L is plotted in Extended Data Fig. 4a. Throughout the following, we take a forward simulation approach, drawing random samples from this distribution. All estimates are based on 100,000 random simulations.

We use this formula to estimate the probability of observing 244 local cases in Miami-Dade county alongside 322 travel-related cases. We approach this by sampling 322 introduction events from L and calculating the total number of local cases in the resulting outbreak (Extended Data Fig. 4b). We also

calculate the probability of observing 244 or fewer local cases in the total outbreak (Extended Data Fig. 4c), finding that R_0 must be 0.6 or smaller to have at least a 1% chance of observing 244 or fewer local cases. As a sensitivity analysis, we additionally model introductions with the assumption that only 50% of travelers are infectious at time of arrival into Miami-Dade County.

We further use this formula to address the probability of observing 3 distinct genetic clusters (F1, F2 and F3) representing 3 introduction events in a sample of 27 ZIKV genomes from Miami-Dade county. We approach this by sampling introduction events until we accumulate 244 local cases according to L, arriving at N introduction events with case counts (j_1, j_2, … j_N). We then sample 27 cases *without replacement* from (j_1, j_2, … j_N) following a hypergeometric distribution and record the number of distinct clusters drawn in the sample. We find that higher values of R_0 result in fewer distinct clusters within the sample of 27 genomes (Extended Data Fig. 4d). We additionally calculate the probability of sampling 3 or fewer distinct genetic clusters in 27 genomes (Extended Data Fig. 4e), finding that R_0 must be greater than 0.6 to have at least a 1% chance of observing 3 or fewer distinct genetic clusters. Additionally, as a sensitivity analysis we model a preferential sampling process in which larger clusters are more likely to be drawn from than smaller clusters. Here, we use a parameter α that enriches the hypergeometric distribution following (j_1^{α}, j_2^{α}, … j_N^{α}).

Under strict assumptions (100% of travelers infectious and random sampling), we find R_0 of approximately 0.6 to be most consistent with both data sources. Under relaxed assumptions (50% of travelers infectious and preferential sampling), we find R_0 of 0.4–0.9 are loosely compatible with the data (Extended Data Fig. 4).

We additionally perform birth-death stochastic simulations assuming a serial interval with mean 20 days[12]. We record the number of stochastic simulations still persisting after a particular number of days for different values of R_0 (Fig. 2c).

Incidence and attack rates

Weekly suspected and confirmed ZIKV case counts from countries and territories within the Americas with local transmission (January 1st to September 18th, 2016) were obtained from the Pan American Health Organization (PAHO)[66]. In most cases, the weekly case numbers per country were only reported in bar graphs, therefore we used WebPlotDigitizer v3.10 (http://arohatgi.info/WebPlotDigitizer) to estimate the numbers. Country and territory total population sizes to calculate weekly and monthly ZIKV incidence rates were also obtained from PAHO[67]. Incidence rates calculated from countries and territories in the Americas during January to June, 2016 (based on the earliest introduction time estimates until the first known cases) were used an estimate for infection likelihood to investigate sources of ZIKV introductions.

To remove any potential reporting biases with incidence rates, we also used ZIKV attack rates (*i.e.*, proportion infected before epidemic burnout) to estimate infection likelihood. Attack rates were calculated using a susceptible–infected–recovered (SIR) transmission model derived from seroprevalence studies and environmental factors as described[68]. Using attack rates as an estimate of infection likelihood, we predict that ~60% of the infected travelers entering Miami came from the Caribbean (Extended Data 7b), which is in agreement with our methods using incidence rates (Fig. 3c). A list of countries and territories used in these analysis can be found in Supplementary Table 2.

Airline and cruise ship traffic

To investigate whether the transmission of ZIKV in Florida coincides with travel patterns from ZIKV endemic regions, we obtained the number of passengers arriving at airports in Florida via commercial air travel. We collated flight data from countries and territories in the Americas with local ZIKV transmission between January and June, 2016 (based on the earliest introduction time estimates until the

first known cases, Supplementary Table 2), arriving at all commercial airports in Florida. The data were obtained from the International Air Transportation Association, which collects data on an estimated 90% of all passenger trips worldwide. We used previously reported flight data from October 2014 through September 2015 to compare traffic patterns across major United States airports[23].

Schedules for cruise ships visiting Miami, Port Canaveral, Port Everglades, Fort Lauderdale, Key West, Jacksonville (all in Florida), Houston, Galveston (both in Texas), Charleston (South Carolina) and New Orleans (Louisiana) ports in the year 2016 were collated from www.cruisett.com and confirmed by cross-referencing ship logs reported by Port of Miami and reported ship schedules. Cruise ship capacities were extracted from www.cruisemapper.com. Every country/territory with ZIKV transmission visited by a cruise ship 10 days (the approximate mean time to ZIKV clearance in human blood [*i.e.*, the infectious period])[69] prior to arrival to Port of Miami was counted as contributing the ship's capacity worth of passengers to Miami to the month of arrival (Supplementary Table 2).

Expected number of travelers infected with Zika virus

We estimated the expected number of travelers entering Miami who were infected with ZIKV (λ) by using the total travel capacity (C) and the likelihood of ZIKV infection (infections (I) per person (N)) from each country/territory (i):

We summed the number of expected infected travelers from each country/territory with ZIKV transmission by region and travel method (flights or cruises). We reported the data as the relative proportion of infected travelers from each region (Fig. 3c, Extended Data Fig. 7a) and as the absolute number of infected travelers (Fig. 3d, Extended Data Fig. 7b, Supplementary Table 2).

Figure Legends

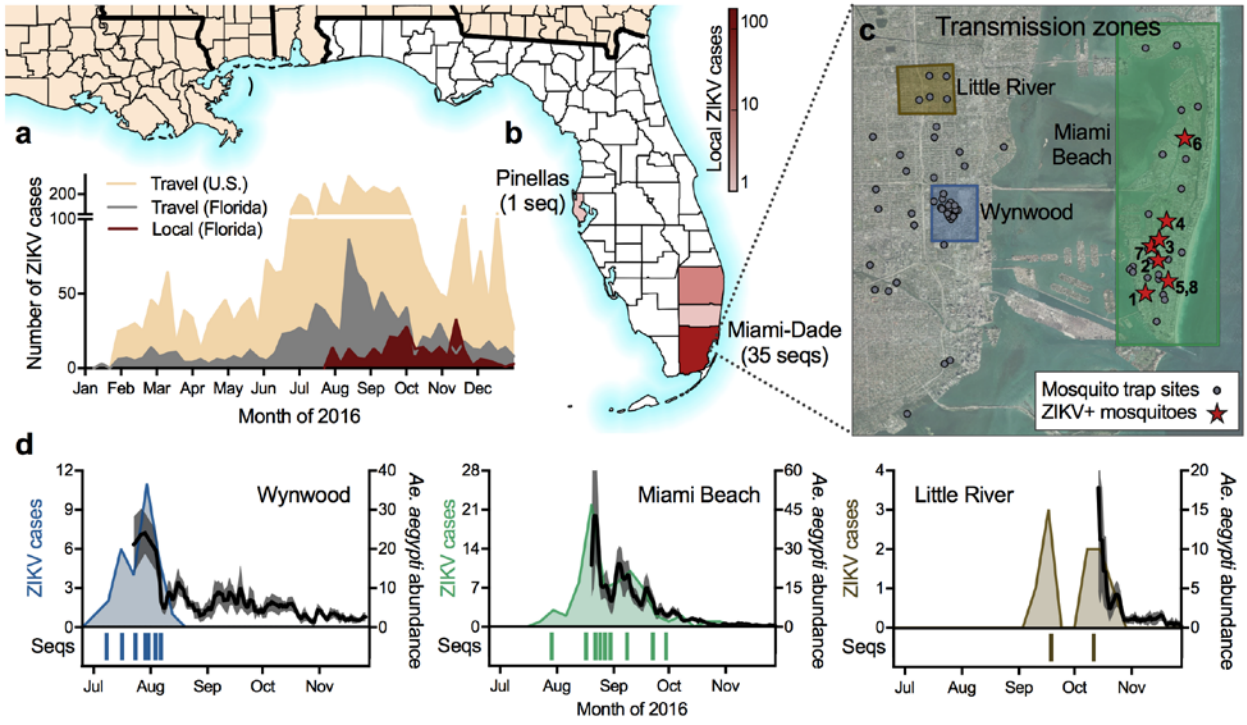


Figure 1 | Zika virus outbreak in Florida. (**a**) Weekly counts of confirmed travel-associated and locally acquired ZIKV cases in 2016. (**b**) Four counties reported probable locally-acquired ZIKV cases in 2016: Miami-Dade (243), Broward (1), Palm Beach (5), Pinellas (1), and undetermined (6). Sequence data (seqs) were generated from the only case reported in Pinellas County and from 28 human cases and 7 *Ae. aegypti* samples from Miami-Dade County. (**c**) The locations of mosquito traps and collected *Ae. aegypti* mosquitoes found to contain ZIKV RNA (ZIKV+) in relation to the transmission zones within Miami. Additional trap sites outside of this area are not shown. (**d**) Temporal distribution of weekly ZIKV cases (top, left y-axis), sequenced cases (bottom), and *Ae. aegypti* abundance per trap night (right y-axis) associated with the three described transmission zones. ZIKV cases and sequences are plotted in relation to onset dates. Sequenced cases without onset dates are not shown. Human cases and *Ae. aegypt*i abundance per week were positively correlated (Spearman r = 0.61, Extended Data Fig. 1e).

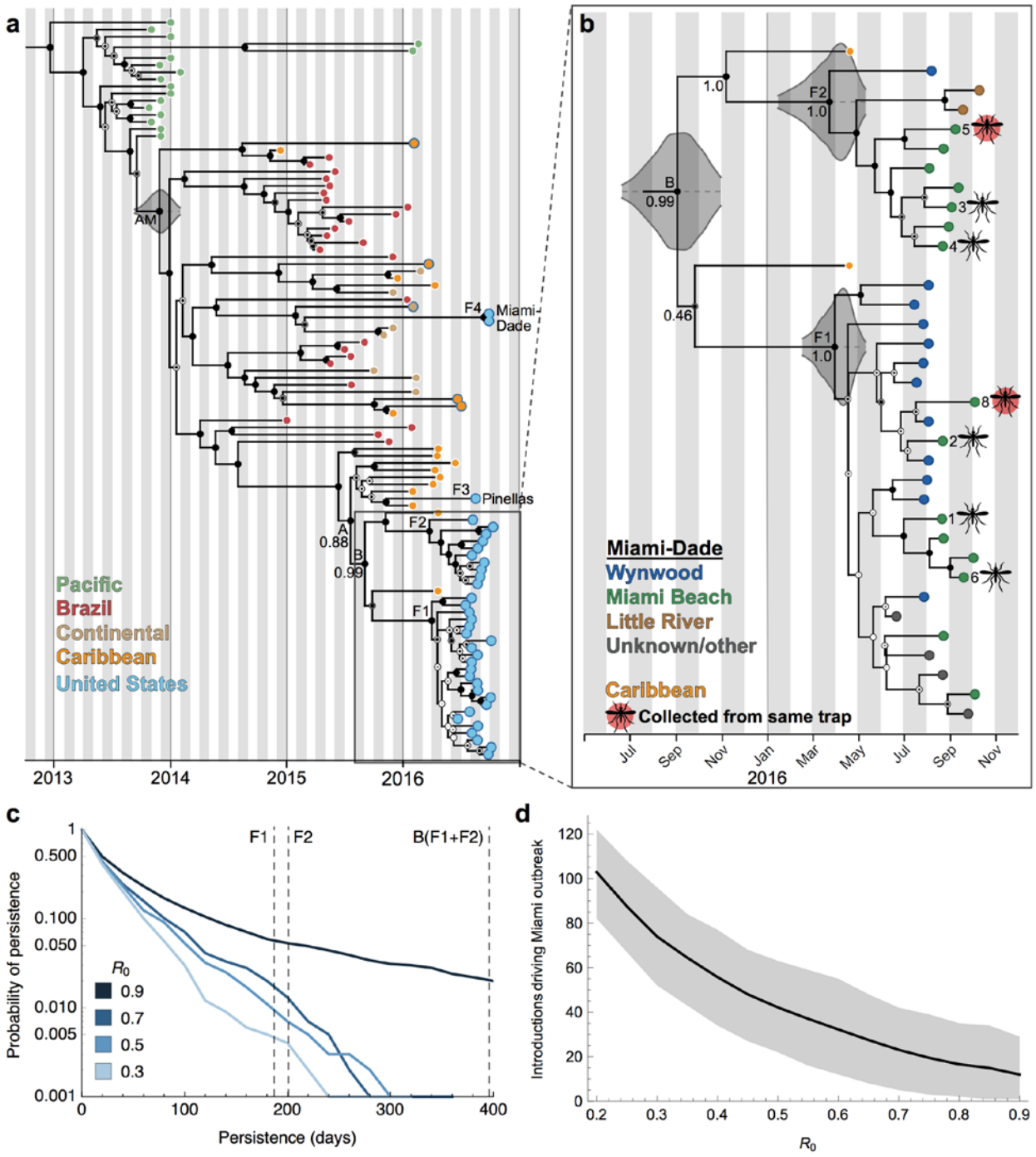


Figure 2 | Multiple introductions of Zika virus into Florida. (**a**) Maximum clade credibility (MCC) tree of ZIKV genomes sequenced from outbreaks in the Pacific islands and the epidemic in the Americas. Tips are colored based on collection location. The five tips outlined in blue but filled with a different color indicates ZIKV cases in the United States associated with travel (fill color indicates the probable origin). Clade posterior probabilities are indicated with black circles inside white circles with posterior probability of 1 filling the entire white circle. The grey violin plot indicates the 95% highest posterior density (HPD) interval for tMRCA for the epidemic in the Americas (AM)[5]. Lineage F4 contains two identical ZIKV genomes from the same patient. (**b**) A zoomed in version of the whole MCC tree showing

the collection locations of Miami-Dade sequences and whether they were sequenced from mosquitoes (numbers correspond to Fig. 1c). 95% HPD intervals are shown for the tMRCAs (**c**) The probability of ZIKV persistence after introduction for different R_0. Persistence is measured as the number of days from initial introduction of viral lineages until their extinction. Vertical dashed lines show the inferred persistence time for lineages F1, F2 and B based on their tMRCA. (**d**) Total number of introductions (mean with 95% CI) that contributed to the outbreak of 244 local cases in Miami-Dade County for different R_0.

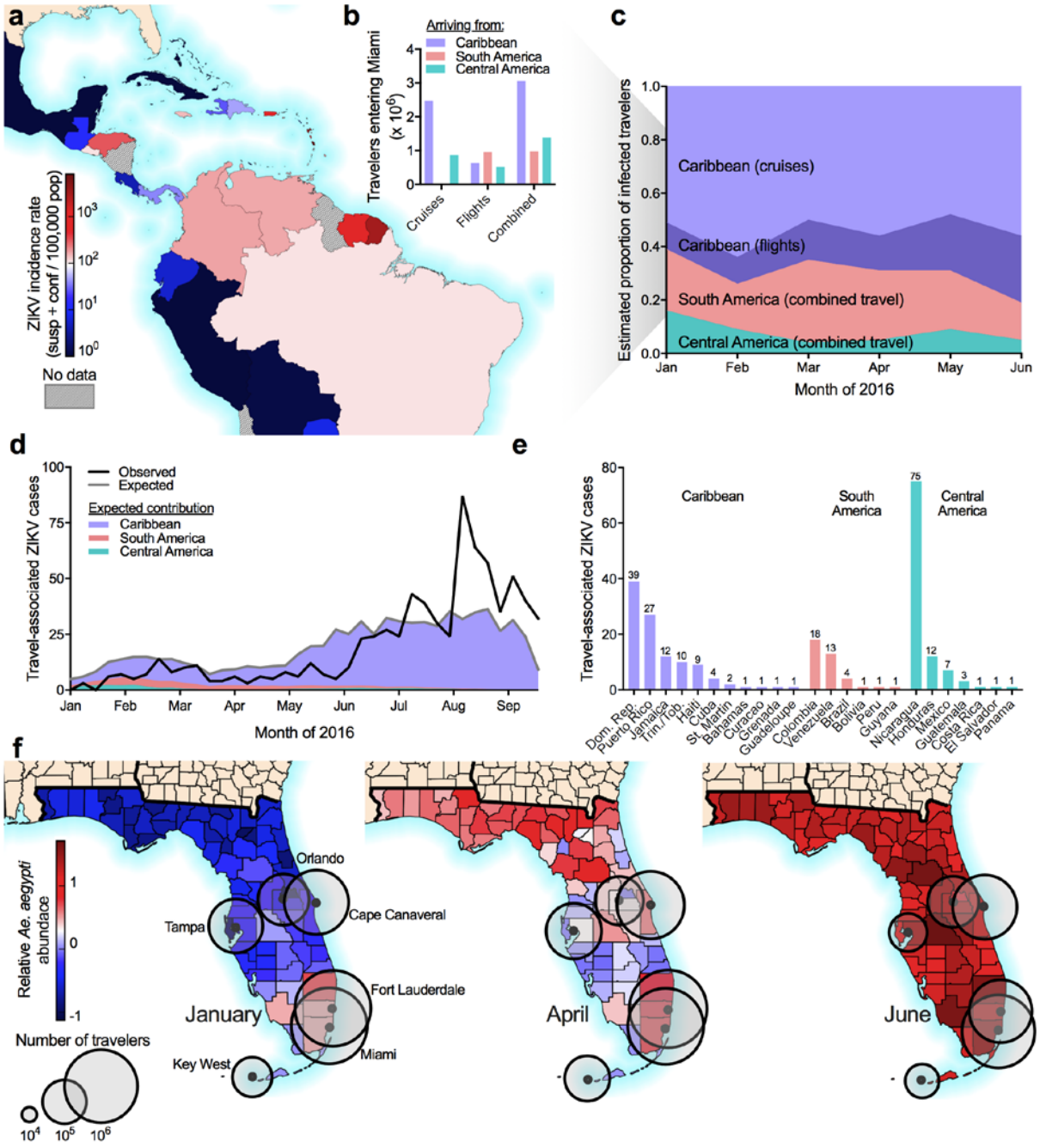


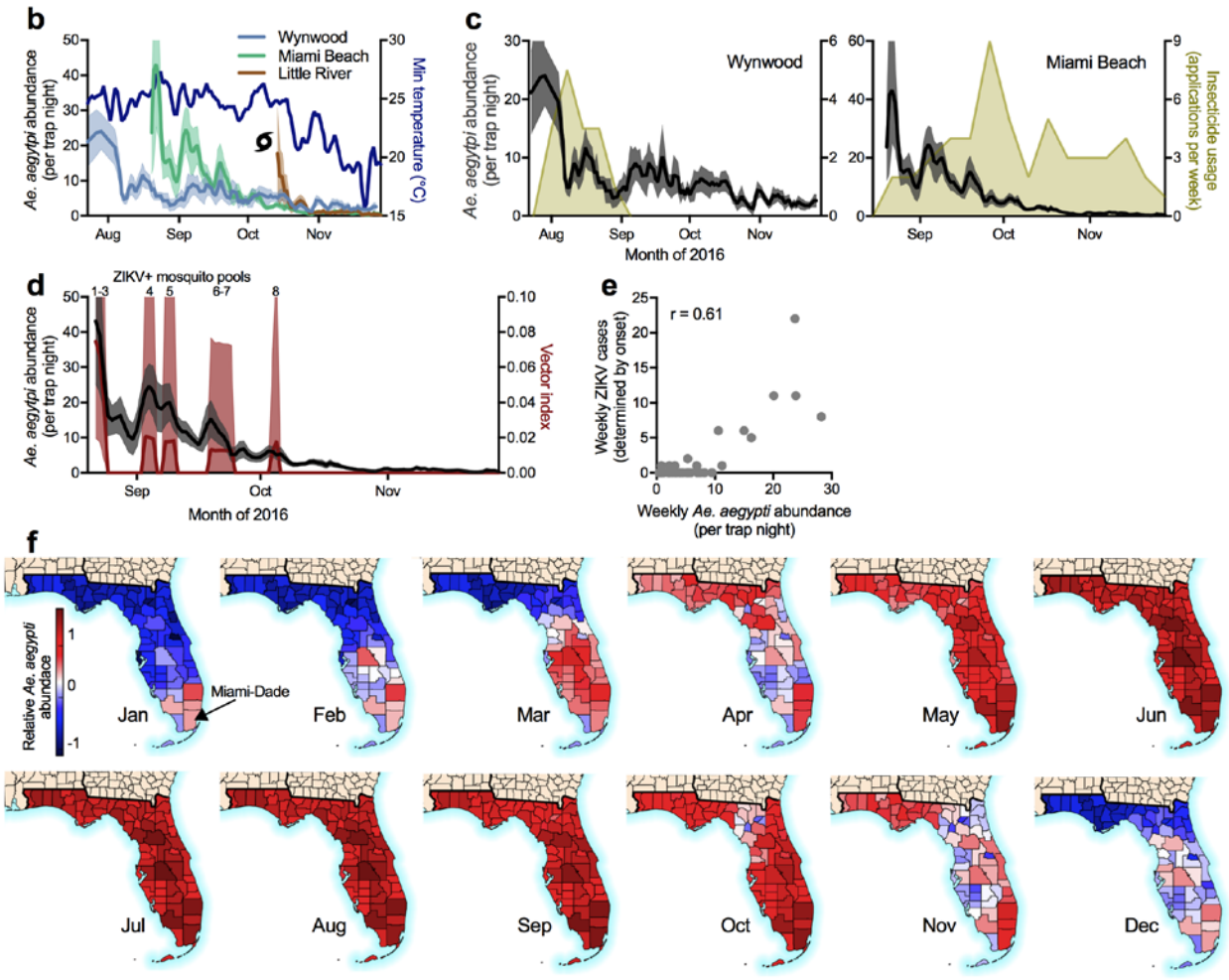
Figure 3 | High probability of Zika virus introductions into Miami from the Caribbean. (**a**) Reported ZIKV cases from January to June, 2016 normalized by total population. (**b**) The number of expected travelers entering Miami during January to June, 2016 by method of travel. (**c**) The number of travelers and the reported ZIKV incidence rate from the country/territory visited were used to estimate the proportion of infected travelers coming from each region with ZIKV in the Americas. (**d**) The observed number of weekly travel-associated ZIKV cases in Florida were plotted with the expected number of ZIKV-infected travelers (as estimated in panel c) coming from all of the Americas (grey line) and the regional contributions (colored areas). (**e**) The countries visited by 236 travel-associated ZIKV cases reported from Miami-Dade county. (**f**) The number of travelers per month (circles) entering Florida cities

via flights and cruises were plotted with relative *Ae. aegypti* abundance. Only ports receiving >10,000 passengers per month are shown.

Extended Data

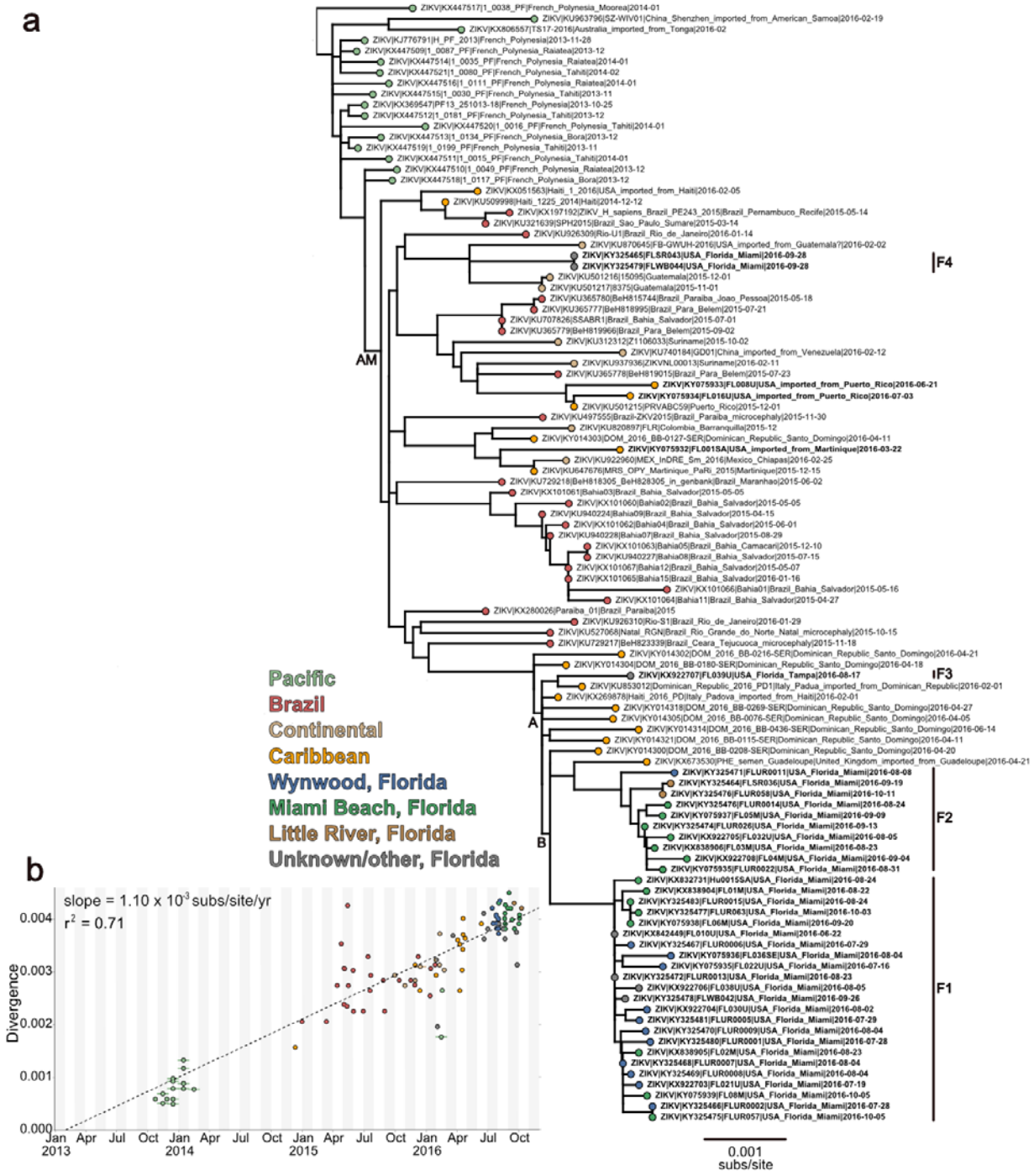
a Miami-Dade mosquito surveillance

Location	Start date	End date	Trap sites	Trap nights	*Ae. aegypti* collected	No. per trap night	Pools tested	ZIKV+ pools	MLE infection rate
Wynwood	Jul 23	Nov 25	29	824	4,972	6.0	738	0	0
Miami Beach	Aug 21	Nov 28	29	2,180	13,265	6.1	1,473	8	0.61 (0.28-1.15)
Little River	Oct 15	Nov 28	5	217	485	2.2	136	0	0
Other Miami	Jun 21	Sep 23	52	235	5,584	23.8	249	0	0



Extended Data Fig. 1 | Miami-Dade mosquito surveillance and relative *Aedes aegypti* abundance. (**a**) Mosquito surveillance data reported from June 21 to November 28, 2016 was used to evaluate the risk of ZIKV infection from mosquito-borne transmission in Miami. A total of 24,306 *Ae. aegypti* and 45 *Ae. albopictus* (none from pools containing ZIKV RNA, data not shown in table) were collected. Trap nights are the total number times each trap site was used and the trap locations are shown in Fig. 1d (some "Other Miami" trap sites are not shown). Up to 50 mosquitoes of the same species and trap night were pooled together for ZIKV RNA testing. The infection rates were calculated using a maximum likelihood estimate (MLE). Miami Beach data includes both North and South Miami Beach. (**b**) *Ae. aegypti* abundance per trap night for each transmission zone was calculated using a sliding window of 3 days with a 1 day step to help normalize some of the natural variation in mosquito abundance data. The

minimum temperatures in Miami were plotted to demonstrate that, in general, mosquito abundance decreases with temperature. The influence of Hurricane Matthew, represented by the hurricane symbol on October 7, is difficult to determine because mosquito abundance was already low. (**c**) Insecticide usage, including truck and aerial adulticides and larvacides, by the Miami-Dade Mosquito Control in Wynwood (left) and Miami Beach (right) was overlaid with *Ae. aegypti* abundance per trap night to demonstrate that intense usage of insecticides may have helped to reduce local mosquito populations. (**d**) Vector index, a relative measure of risk of mosquito-borne transmission, is the product of mosquito abundance and the MLE infection rate. Data is shown as the mean with 95% CI using a sliding window of 3 days with a 1 day step. When infected mosquitoes are present, vector index is heavily influenced by mosquito abundance. (**e**) If mosquito abundance influences vector index, and vector index is an estimate of human risk[37], then a correlation between *Ae. aegypti* abundance and human ZIKV cases should be expected. The number of weekly ZIKV cases (based on symptoms onset) was correlated with mean *Ae. aegypti* abundance per trap night determined from the same week and zone (Fig. 1e, Spearman r = 0.61). (**f**) Relative *Ae. aegypti* abundance for each Florida county and month was estimated using a multivariate regression model, demonstrating spatial and temporal heterogeneity for the risk of ZIKV infection.

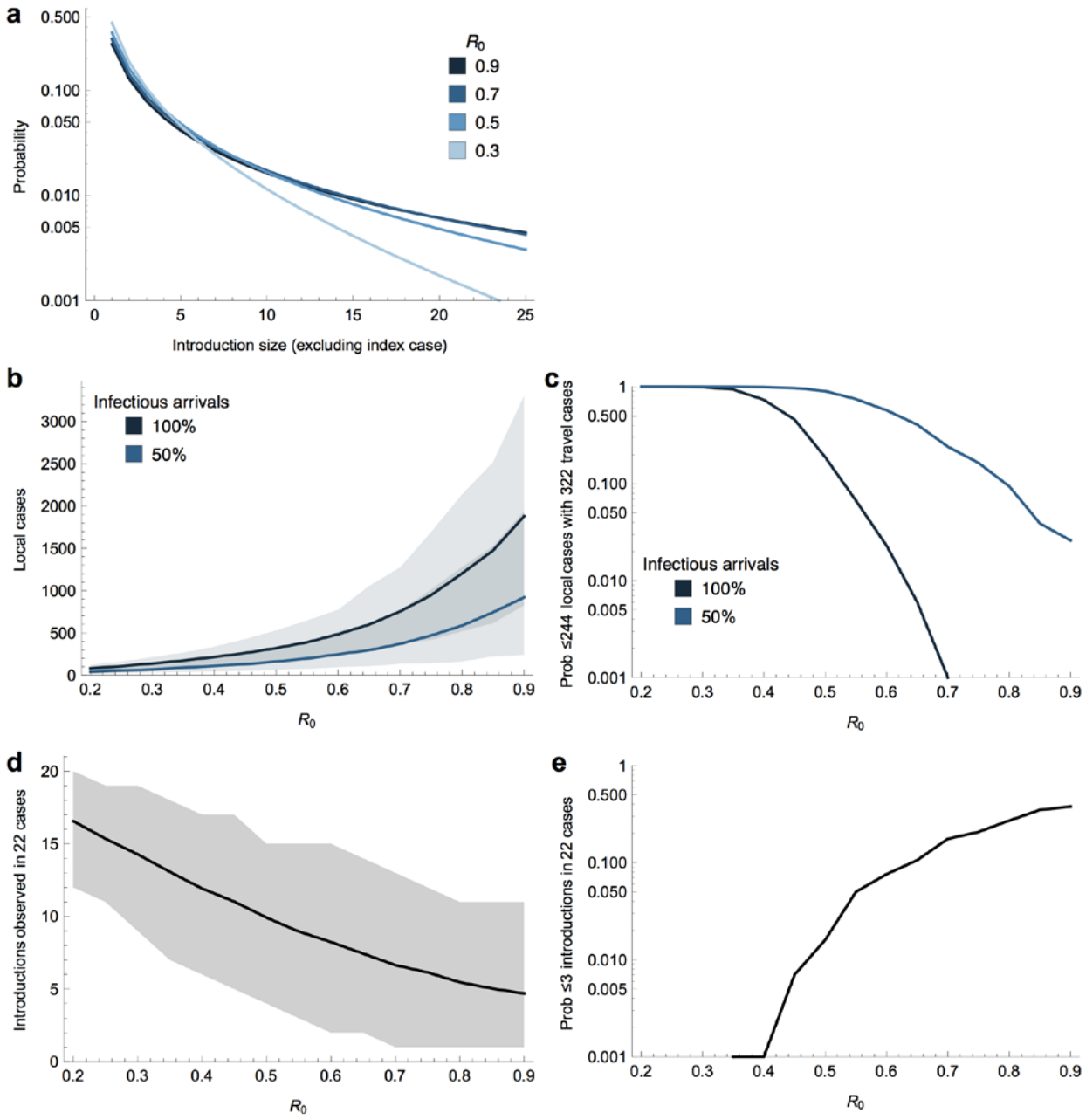


Extended Data Fig. 2 | Maximum likelihood tree (a) and root-to-tip regression (b) of Zika virus genomes from Pacific islands and the epidemic in Americas. (**a**) Tips are coloured by location, labels in bold indicate samples sequenced in this study, Florida clusters F1-4 are indicated by vertical lines to the right of the tree. (**b**) Linear regression of tip dates against divergence from root based on sequences with known collection dates indicates that the evolutionary rate estimate for the entire ZIKV phylogeny is 1.10×10^{-3} nucleotide substitutions/site/year (subs/site/yr). This is consistent with BEAST analyses using a strict molecular clock and a skyline tree prior, the best-performing combination of clock and demographic

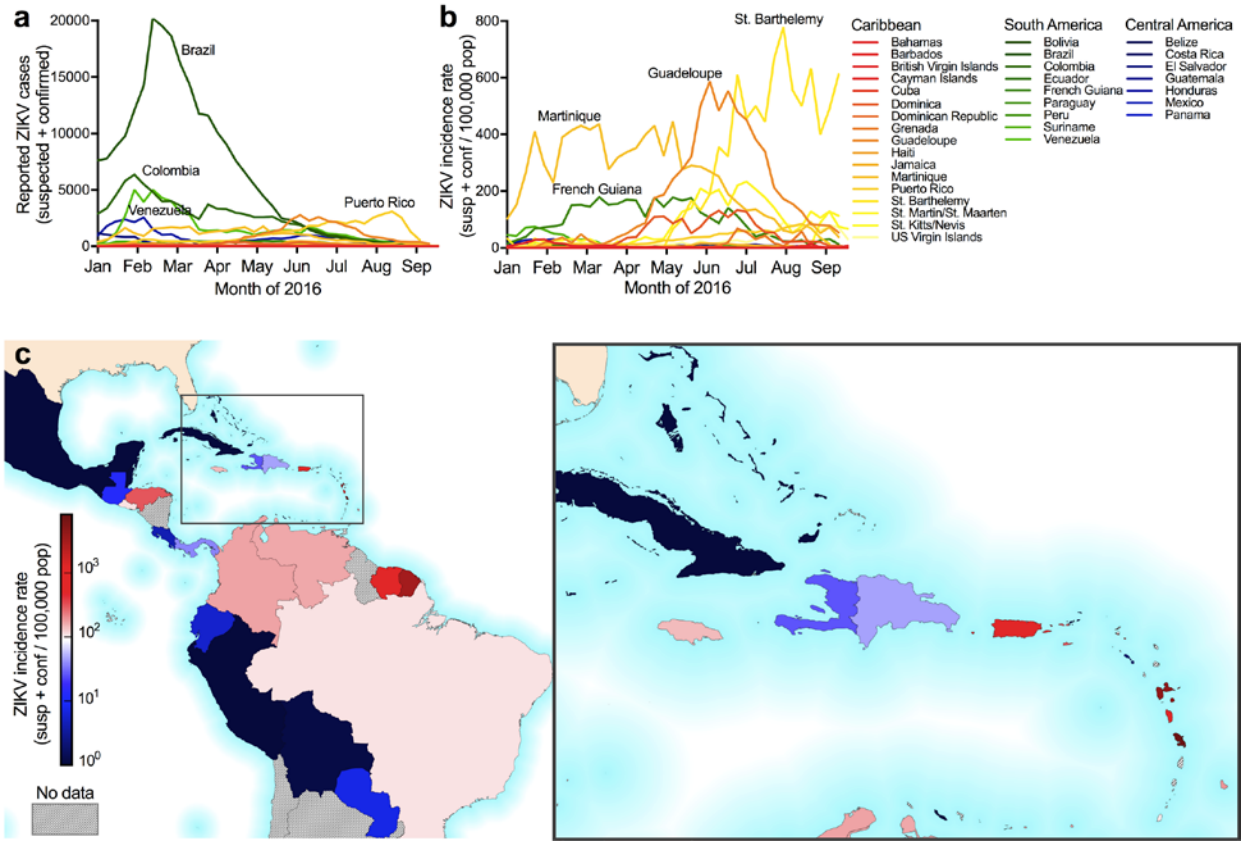
model according to marginal likelihood estimates, indicating that the evolutionary rate is 1.08×10^{-3} (95% highest posterior density: 0.97 - 1.19×10^{-3}) subs/site/yr. These values are in agreement with previous estimates calculated based on ZIKV genomes from Brazil[5].

Extended Data Fig. 3 | Molecular clock dating of Zika virus clades sequenced from Florida. Maximum clade credibility (MCC) tree of ZIKV genomes collected from Pacific islands and the epidemic in Americas. Circles at the tips are colored based on origin location. Clade posterior probabilities are indicated with black circles inside white circles with posterior probability of 1 filling the entire white circle. The grey violin plot indicates the 95% highest posterior density (HPD) interval for the tMRCA of

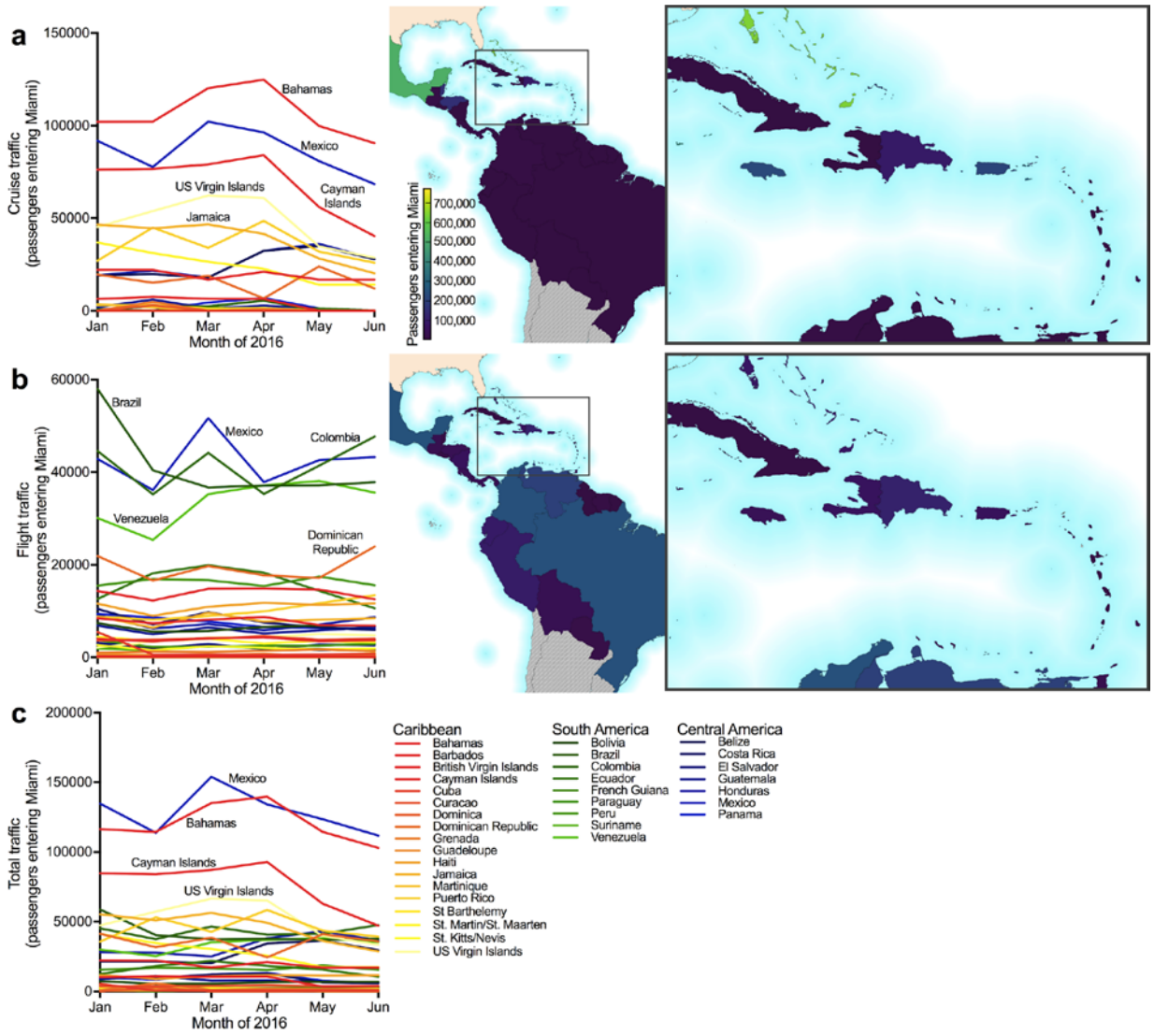
the American epidemic. We estimated that the tMRCA for the ongoing epidemic in the Americas occurred during October, 2013 (node AM, Extended Table 1, 95% BCI: August, 2013-January, 2014), which is consistent with previous analysis based on ZIKV genomes from Brazil[5] (node AM, Fig. 2a).



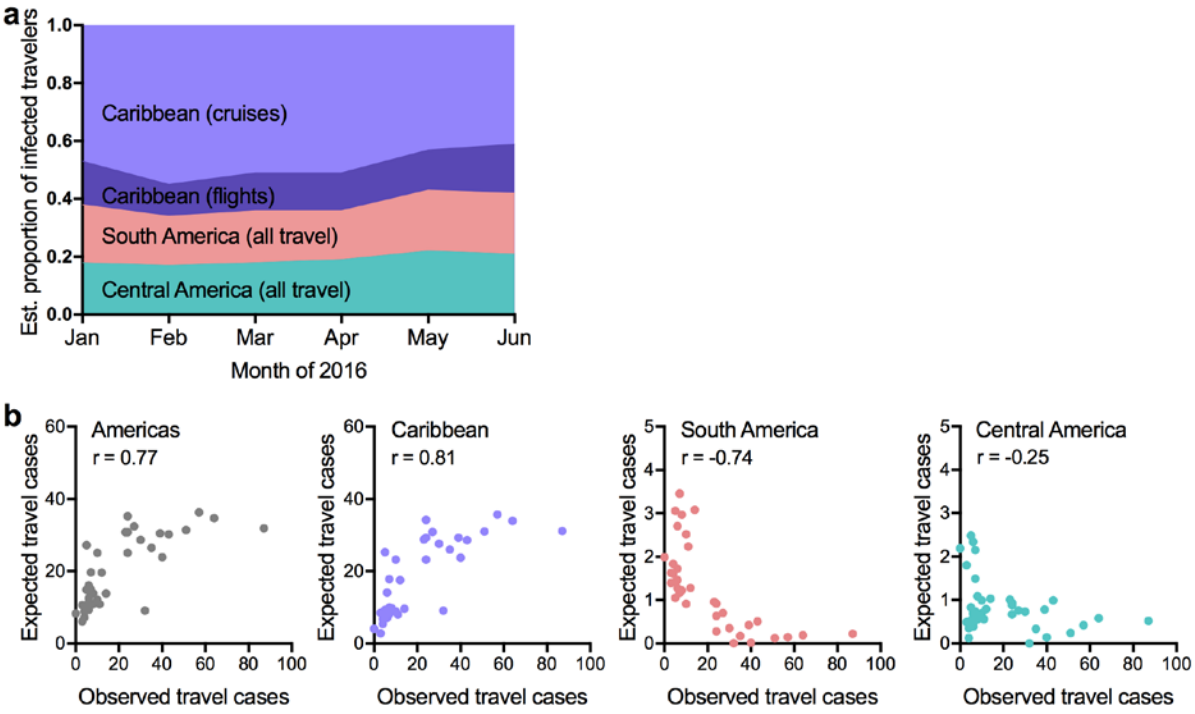
Extended Data Fig. 4 | Estimation of basic reproductive number and number of introductions. (**a**) Probability distribution of total number of cases caused by a single introduction (excluding the index case) for different values of R_0. (**b**) Mean and 95% uncertainty interval for total number of local cases caused by 322 introduction events for different values of R_0. (**c**) Probability of observing 244 of fewer local cases with 322 introduction events for different values of R_0. (**d**) Mean and 95% uncertainty interval for total number of distinct clusters observed in 22 sequenced cases for different values of R_0. (**e**) Probability of observing 3 or fewer clusters in 22 sequenced cases for different values of R_0.



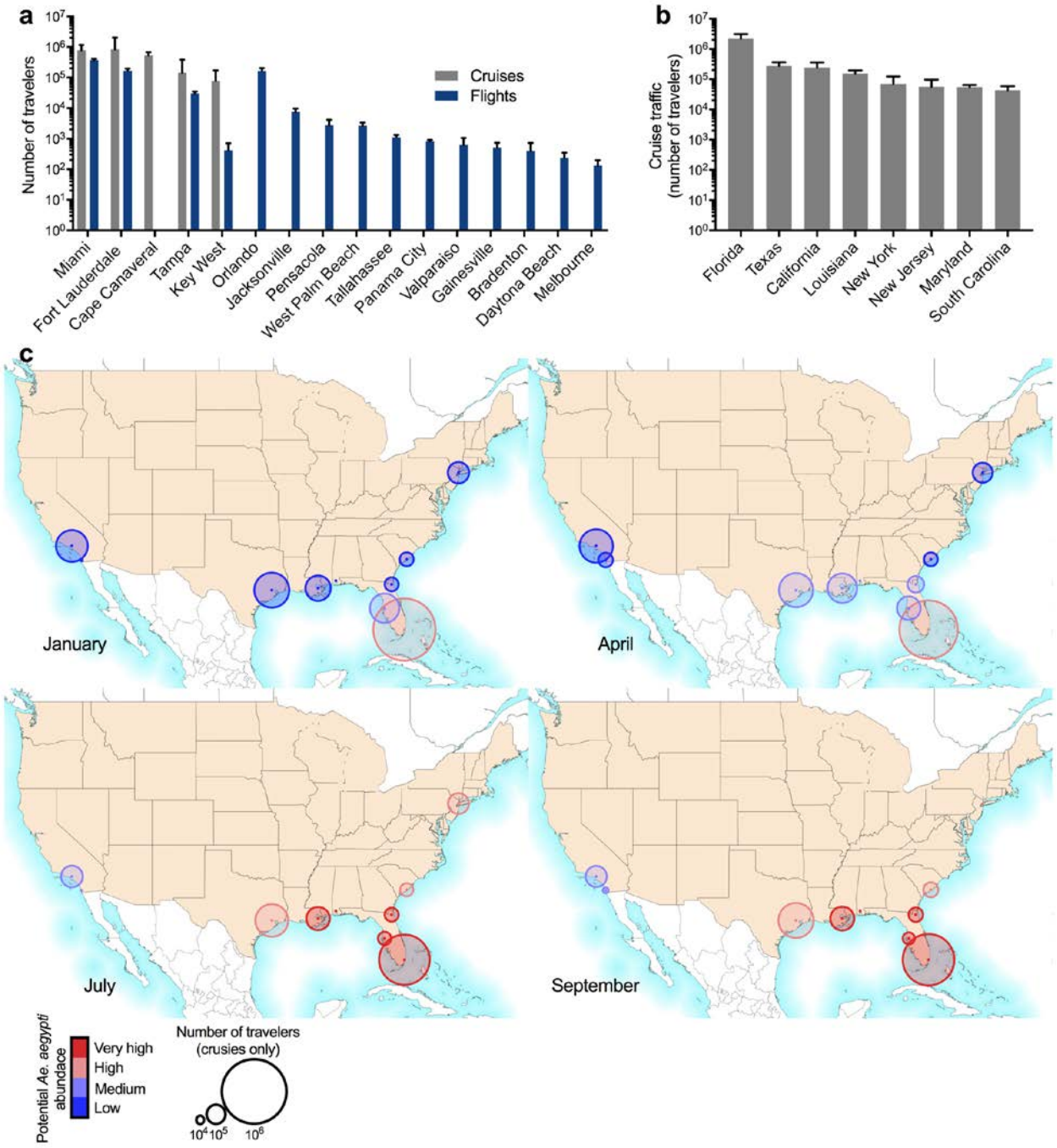
Extended Data Fig. 5 | Weekly reported Zika virus case numbers and incidence rates in the Americas. (**a**) ZIKV cases (suspected and confirmed) and (**b**) incidence rates (normalized per 100,000 population) are shown for each country or territory with available data per epidemiological week from January 1st to September 18th, 2016. (**c**) Each country or territory with available data is colored by its reported ZIKV incidence rate from January to June, 2016 (the time frame for analysis of ZIKV introductions into Florida).



Extended Data Fig. 6 | Cruise and flight traffic entering Miami from regions with Zika virus transmission. The number of passengers entering Miami, by either (**a**) cruises or (**b**) flights, from each country or territory in the Americas with ZIKV transmission per month (left panel). The center map and inset show the cumulative numbers of travelers entering Miami during January to June, 2016 (the time frame for analysis of ZIKV introductions into Florida) from each country or territory per method of travel. (**c**) The total traffic (*i.e.* cruises and flights) is shown entering Miami per month. (**d**) Summarizes the contribution of each travel method towards the cumulative numbers of travelers entering Miami from the Caribbean, South America, and Central America during January to June, 2016.



Extended Data Fig. 7 | Expected number of Zika virus infected travelers from the Caribbean is correlated with the total observed number of travel infections. (**a**) The ZIKV incidence rates used in the study may have been biased by reporting accuracy. Therefore, we alternately used projected ZIKV attack rates[68] (*i.e.* predicted proportion of population infected before epidemic burnout) along with travel capacities to estimate the proportion of infected travelers entering Miami from each region with ZIKV in the Americas. About 60% of the infected travelers are expected to have arrived from the Caribbean, similar to our results using incidence rates (Fig. 3c). (**b**) The expected number of travel-associated ZIKV cases were estimated by the number of travelers coming into Miami from each country/territory (travel capacity) and the in-country/territory infection likelihood (incidence rate per person) per week. The expected travel cases were summed from all of the Americas (left), Caribbean (left center), South America (right center), and Central America (right) and plotted with the observed travel-associated ZIKV cases. Numbers in each plot indicate Spearman correlation coefficients.



Extended Data Fig. 8 | High potential for Zika virus introductions into Miami via cruises. (**a**) The monthly average (95% CI) number of expected travelers (*i.e.* capacity) entering all Florida ports from countries/territories in the Americas with ZIKV transmission during January to June, 2016. (**b**) The monthly average (95% CI) number of travelers entering the major United States ports via cruises from countries/territories in the Americas with ZIKV transmission during January to June, 2016. Ports were combined within states. (**c**) The monthly cruise capacity for each major United States port (shown as circle diameter) with monthly potential *Ae. aegypti* abundance (circle color), as previously estimated[18]. Cruise capacities from Houston and Galveston, TX were combined.

Extended Data Table 1 | Time of the most recent common ancestor.

Model combination	Clade AM tMRCA			Clade A tMRCA			Clade B tMRCA		
	Mean	Lower 95% BCI	Upper 95% BCI	Mean	Lower 95% BCI	Upper 95% BCI	Mean	Lower 95% BCI	Upper 95% BCI
Strict, Constant	2013.86	2013.66	2014.06	2015.57	2015.39	2015.73	2015.68	2015.49	2015.84
Strict, Exponential	2013.88	2013.66	2014.06	2015.58	2015.40	2015.72	2015.68	2015.52	2015.84
Strict, Bayesian SkyGrid	2013.85	2013.64	2014.05	2015.57	2015.39	2015.72	2015.68	2015.50	2015.83
Strict, Bayesian Skyline	2013.90	2013.69	2014.08	2015.54	2015.38	2015.73	2015.67	2015.47	2015.83
UCLN, Constant	2013.90	2013.66	2014.12	2015.62	2015.41	2015.80	2015.73	2015.53	2015.91
UCLN, Exponential	2013.90	2013.65	2014.11	2015.62	2015.42	2015.78	2015.73	2015.54	2015.90
UCLN, Bayesian SkyGrid	2013.90	2013.66	2014.12	2015.62	2015.42	2015.80	2015.73	2015.54	2015.92
UCLN, Bayesian Skyline	2013.93	2013.71	2014.15	2015.56	2015.38	2015.77	2015.70	2015.50	2015.89

Model combination	Clade F1 tMRCA			Clade F2 tMRCA		
	Mean	Lower 95% BCI	Upper 95% BCI	Mean	Lower 95% BCI	Upper 95% BCI
Strict, Constant	2016.14	2016.00	2016.26	2016.15	2015.97	2016.29
Strict, Exponential	2016.14	2016.00	2016.26	2016.15	2015.99	2016.30
Strict, Bayesian SkyGrid	2016.14	2016.00	2016.26	2016.15	2015.97	2016.29
Strict, Bayesian Skyline	2016.25	2016.13	2016.36	2016.23	2016.04	2016.36
UCLN, Constant	2016.16	2016.03	2016.29	2016.17	2015.99	2016.32
UCLN, Exponential	2016.16	2016.02	2016.28	2016.17	2016.00	2016.31
UCLN, Bayesian SkyGrid	2016.16	2016.01	2016.28	2016.18	2016.00	2016.32
UCLN, Bayesian Skyline	2016.26	2016.13	2016.37	2016.24	2016.04	2016.38

Extended Data Table 2 | Validation of sequencing results.

Sample	Amplicon method	NGS platform	Mismatches/nucleotides covered[a]		
			3' UTR	CDS	5' UTR
FL01M	35 × 400 bp	Ion S5	1/80	0/10272	7/252
	75 × ~200 bp[b]	Ion S5	2/75	0/10272	4/205
	5 × ~2,200 bp	MiSeq	0/80	0/10272	0/32
FL03M	35 × 400 bp	Ion S5	3/87	0/10272	20/252
	75 × ~200 bp[b]	Ion S5	4/78	0/10272	5/198
	5 × ~2,200 bp	MiSeq	0/82	0/10272	0/32

[a] Compared to the consensus genomes generated by sequencing 35 × 400 bp amplicons on the MiSeq.
[b] Amplicons produced using Ion AmpliSeq and 875 custom ZIKV primers.
NGS, next-generation sequencing; UTR, untranslated region; CDS, coding sequence.

Extended Data Table 3 | Model selection to infer time-structured phylogenies.

Model combination	Path Sampling	Ranking	Stepping Stone	Ranking
Strict, Constant	-24305.105	3	-24307.152	3
Strict, Exponential	-24310.130	4	-24312.460	4
Strict, Bayesian SkyGrid	-24301.305	2	-24303.142	2
Strict, Bayesian Skyline	-24287.785	**1**	-24290.762	**1**
UCLN, Constant	-25467.164	6	-25470.141	6
UCLN, Exponential	-25468.457	7	-25471.066	7
UCLN, Bayesian SkyGrid	-25471.792	8	-25474.032	8
UCLN, Bayesian Skyline	-25447.942	5	-25450.574	5

Supplemental Data

Supplementary Table 1. Summary of the Zika virus sequencing data produced in this study.

Supplementary Table 2. Epidemiological data and travelers entering Miami, Florida from January to June, 2016.

Supplementary Table 3. Probe sequences used for RNA Access targeted enrichment of Zika virus.

References

1. WHO | Zika virus and complications. (2016).
2. Lazear, H. M. & Diamond, M. S. Zika Virus: New Clinical Syndromes and Its Emergence in the Western Hemisphere. *J. Virol.* **90,** 4864–4875 (2016).
3. Likos, A. *et al.* Local Mosquito-Borne Transmission of Zika Virus - Miami-Dade and Broward Counties, Florida, June-August 2016. *MMWR Morb. Mortal. Wkly. Rep.* **65,** 1032–1038 (2016).
4. Hennessey, M., Fischer, M. & Staples, J. E. Zika Virus Spreads to New Areas — Region of the Americas, May 2015–January 2016. *MMWR Morb. Mortal. Wkly. Rep.* **65,** 1–4 (2016).
5. Faria, N. R. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352,** 345–349 (2016).
6. Regional Zika Epidemiological Update (Americas). *Pan American Health Organization* (2016). Available at: http://www.paho.org/hq/index.php?option=com_content&view=article&id=11599&Itemid=41691&lang=en.
7. Weger-Lucarelli, J. *et al.* Vector Competence of American Mosquitoes for Three Strains of Zika Virus. *PLoS Negl. Trop. Dis.* **10,** e0005101 (2016).
8. Guerbois, M. *et al.* Outbreak of Zika virus infection, Chiapas State, Mexico, 2015, and first confirmed transmission by Aedes aegypti mosquitoes in the Americas. *J. Infect. Dis.* (2016).
9. Ferreira-de-Brito, A. *et al.* First detection of natural infection of Aedes aegypti with Zika virus in Brazil and throughout South America. *Mem. Inst. Oswaldo Cruz* 0 (2016).
10. Chouin-Carneiro, T. *et al.* Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. *PLoS Negl. Trop. Dis.* **10,** e0004543 (2016).
11. Kraemer, M. U. G. *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *Elife* **4,** e08347 (2015).
12. Ferguson, N. M. *et al.* Countering the Zika epidemic in Latin America. *Science* **353,** 353–354

(2016).

13. Mosquito-Borne Disease Surveillance. *Florida Department of Health* Available at: http://www.floridahealth.gov/diseases-and-conditions/mosquito-borne-diseases/surveillance.html.
14. Teets, F. D. *et al.* Origin of the dengue virus outbreak in Martin County, Florida, USA 2013. *Virol Rep* **1-2,** 2–8 (2014).
15. Graham, A. S. *et al.* Mosquito-associated dengue virus, Key West, Florida, USA, 2010. *Emerg. Infect. Dis.* **17,** 2074–2075 (2011).
16. Kendrick, K. *et al.* Notes from the field: transmission of chikungunya virus in the continental United States—Florida, 2014. *MMWR Morb. Mortal. Wkly. Rep.* **63,** 1137 (2014).
17. Adalja, A. A., Kirk Sell, T., Bouri, N. & Franco, C. Lessons Learned during Dengue Outbreaks in the United States, 2001–2011. *Emerging Infectious Disease journal* **18,** 608 (2012).
18. Monaghan, A. J. *et al.* On the Seasonal Occurrence and Abundance of the Zika Virus Vector Mosquito Aedes Aegypti in the Contiguous United States. *PLoS Curr.* **8,** (2016).
19. Dzul-Manzanilla, F. *et al.* Evidence of vertical transmission and co-circulation of chikungunya and dengue viruses in field populations of Aedes aegypti (L.) from Guerrero, Mexico. *Trans. R. Soc. Trop. Med. Hyg.* **110,** 141–144 (2016).
20. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *bioRxiv* 098913 (2017).
21. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).
22. Dinh, L., Chowell, G., Mizumoto, K. & Nishiura, H. Estimating the subcritical transmissibility of the Zika outbreak in the State of Florida, USA, 2016. *Theor. Biol. Med. Model.* **13,** 20 (2016).
23. Nelson, B. *et al.* Travel Volume to the United States from Countries and U.S. Territories with Local Zika Virus Transmission. *PLoS Curr.* **8,** (2016).
24. Robert, M. A. *et al.* Modeling Mosquito-Borne Disease Spread in U.S. Urbanized Areas: The Case of Dengue in Miami. *PLoS One* **11,** e0161365 (2016).

25. Ramos, M. M. *et al.* Epidemic Dengue and Dengue Hemorrhagic Fever at the Texas–Mexico Border: Results of a Household-based Seroepidemiologic Survey, December 2005. *Am. J. Trop. Med. Hyg.* **78,** 364–369 (2008).
26. McCarthy, M. First US case of Zika virus infection is identified in Texas. *BMJ* **352,** i212 (2016).
27. Duffy, M. R. *et al.* Zika virus outbreak on Yap Island, Federated States of Micronesia. *N. Engl. J. Med.* **360,** 2536–2543 (2009).
28. Semenza, J. C. *et al.* International dispersal of dengue through air travel: importation risk for Europe. *PLoS Negl. Trop. Dis.* **8,** e3278 (2014).
29. Zika virus. *Florida Department of Health* Available at: http://www.floridahealth.gov/diseases-and-conditions/zika-virus/index.html?utm_source=flhealthIndex.
30. Daily Zika update. *Florida Department of Health* Available at: http://www.floridahealth.gov/newsroom/all-articles.html.
31. Zika virus case counts in the US. *The Centers for Disease Control and Prevention* (2016). Available at: http://www.cdc.gov/zika/geo/united-states.html.
32. Rabe, I. B. *et al.* Interim Guidance for Interpretation of Zika Virus Antibody Test Results. *MMWR Morb. Mortal. Wkly. Rep.* **65,** 543–546 (2016).
33. Waggoner, J. J. & Pinsky, B. A. Zika Virus: Diagnostics for an Emerging Pandemic Threat. *J. Clin. Microbiol.* **54,** 860–867 (2016).
34. Lanciotti, R. S. *et al.* Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerg. Infect. Dis.* **14,** 1232–1239 (2008).
35. Interim Guidance for Zika Virus Testing of Urine - United States, 2016. *MMWR Morb. Mortal. Wkly. Rep.* **65,** 474 (2016).
36. Biggerstaff, B. PooledInfRate, version 4.0. *An Excel Add-In to Compute Infection Rates from Pooled Data. Centers for Disease Control, Fort Collins, CO* (2009).
37. Fauver, J. R. *et al.* Temporal and Spatial Variability of Entomological Risk Indices for West Nile Virus Infection in Northern Colorado: 2006-2013. *J. Med. Entomol.* **53,** 425–434 (2016).

38. National Centers for Environmental Information. *National Oceanic and Atmospheric Administration* Available at: https://www.ncdc.noaa.gov/.

39. Bogoch, I. I. *et al.* Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study. *Lancet Infect. Dis.* (2016). doi:10.1016/S1473-3099(16)30270-5

40. Hwang, W.-H. & He, F. Estimating abundance from presence/absence maps. *Methods Ecol. Evol.* **2,** 550–559 (2011).

41. Corman, V. M. *et al.* Clinical comparison, standardization and optimization of Zika virus molecular detection. *Bull. World Health Organ.* (2016).

42. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

43. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

44. Köster, J. & Rahmann, S. Building and documenting workflows with python-based snakemake. in *OASIcs-OpenAccess Series in Informatics* **26,** (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2012).

45. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).

46. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

47. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27,** 863–864 (2011).

48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

49. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).

50. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).

51. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22,** 160–174 (1985).

52. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39,** 306–314 (1994).

53. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772 (2012).

54. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2,** vew007 (2016).

55. Shapiro, B., Rambaut, A. & Drummond, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23,** 7–9 (2006).

56. Ferreira, M. A. R. & Suchard, M. A. Bayesian analysis of elapsed times in continuous-time Markov chains. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **36,** 355–368 (2008).

57. Baele, G. *et al.* Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29,** 2157–2167 (2012).

58. Xie, W., Lewis, P. O., Fan, Y., Kuo, L. & Chen, M.-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* **60,** 150–160 (2011).

59. Gelman, A. & Meng, X.-L. Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling. *Stat. Sci.* **13,** 163–185 (1998).

60. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22,** 1185–1192 (2005).

61. Churcher, T. S. *et al.* Public health. Measuring the path toward malaria elimination. *Science* **344,** 1230–1232 (2014).

62. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of

individual variation on disease emergence. *Nature* **438,** 355–359 (2005).

63. Nishiura, H., Yan, P., Sleeman, C. K. & Mode, C. J. Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks. *J. Theor. Biol.* **294,** 48–55 (2012).
64. Perkins, T. A., Scott, T. W., Le Menach, A. & Smith, D. L. Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. *PLoS Comput. Biol.* **9,** e1003327 (2013).
65. Kraemer, M. U. G. *et al.* Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *J. R. Soc. Interface* **12,** 20150468 (2015).
66. Zika-Epidemiological Report. *Pan American Health Organization* Available at: http://www.paho.org/hq/index.php?option=com_content&view=article&id=11603&Itemid=41696&lang=en.
67. Zika Cumulative Cases. *Pan American Health Organization* Available at: http://www.paho.org/hq/index.php?option=com_content&view=article&id=12390&Itemid=42090&lang=en.
68. Perkins, T. A., Siraj, A. S., Ruktanonchai, C. W., Kraemer, M. U. G. & Tatem, A. J. Model-based projections of Zika virus infections in childbearing women in the Americas. *Nat Microbiol* **1,** 16126 (2016).
69. Lessler, J. T. *et al.* Times to key events in the course of Zika infection and their implications: a systematic review and pooled analysis. *Bull. World Health Organ.* (2016).

67. Lessler, J. T. *et al.* Times to key events in the course of Zika infection and their implications: a systematic review and pooled analysis. *Bull. World Health Organ.* (2016).